ISRCTN32473131 https://doi.org/10.1186/ISRCTN32473131

Chart review of patients with chronic obstructive pulmonary disease, using medical records and artificial intelligence

Submission date
25/11/2019

Recruitment status
No longer recruiting

Registration date
24/01/2020

Overall study status
Completed

Last Edited
24/01/2020

Condition category
Respiratory

[ ] Prospectively registered
[ ] Protocol
[ ] Statistical analysis plan
[ ] Results
[ ] Individual participant data
[ ] Record updated in last year

Plain English summary of protocol

Background and study aims

Chronic obstructive pulmonary disease (COPD) is the third leading cause of death in the world since 2003. Many people suffer from this disease or its complications for many years and die prematurely. In the European Union, the total direct costs of respiratory diseases are estimated to be around 6% of the total healthcare budget, with COPD accounting for 56% (38.6 billion Euros) of the costs of respiratory diseases.

In the natural history of COPD, many patients may experience acute exacerbations (AECOPD) that are described as episodes of sustained worsening of the respiratory symptoms that result in additional therapy. These episodes of exacerbation that often require being seen in the emergency department and/or a hospital admission are associated with significant morbidity and mortality; they are responsible for a significant portion of the economic burden of the disease too. The pharmacological approach used in the management of AECOPD (inhaled bronchodilators, corticosteroids and antibiotics), has the objective to minimize the negative impact of the current exacerbation and to prevent subsequent events. Despite the collaborative effort between the European Respiratory Society, the American Thoracic Society and others to provide clinical recommendations for the prevention of AECOPD, there is still a considerable number of patients that are prone to suffer from recurrent exacerbations and to experience a more severe impairment in health status. Based on all the above, the aim of this study is to identify the factors potentially associated with hospital admission in patients with AECOPD in English-, French-, German-, and Spanish-speaking countries, and to develop a predictive model that predicts the risk of hospitalization in this group of patients, by using artificial intelligence. In this study the researchers propose to take advantage of SAVANA, a new clinical platform, created in the context of the era of electronic medical records (EMRs), to analyse the information included in the electronic medical files (i.e., big data). This clinical platform is a powerful free-text analysis engine, capable of meaningfully interpreting the contents of the EMRs, regardless of the management system in which they operate. In this context, this machine learning analytical method can be used to build a flexible, customized and automated predictive model using the information available in EMRs.

Who can participate?
Adults both genders with Chronic obstructive pulmonary disease

What does the study involve?
For patients there is no intervention, as the data is extracted from their electronic medical records.

What are the possible benefits and risks of participating?
The benefits is to generate an automated predictive model with the use of machine learning that predicts the risk of hospitalization in patients with AECOPD.

Where is the study run from?
In 80 sites distributed in English, French, German and Spanish speaking countries (UK, Canada, USA, France, Belgium, Switzerland, Germany, Austria, Spain)

When is the study starting and how long is it expected to run for?
April 2019 to December 2020

Who is funding the study?
European Commission with a grant Horizon 2020 on research and innovation, Brussels, Belgium

Who is the main contact?
Prof. Rob Stockley
rob.stockley@uhb.nhs.uk

Contact information

Type(s)
Scientific

Contact name
Prof Robert Stockley

Contact details
Queen Elizabeth Hospital
Mindelsohn Way
Edgbaston
Birmingham
United Kingdom
B9 5SS
+44 (0)121 3716808
rob.stockley@uhb.nhs.uk

Additional identifiers

EudraCT/CTIS number
Nil known

IRAS number

ClinicalTrials.gov number
Nil known

Secondary identifying numbers
BigCOPData

Study information

Scientific Title
Chart review of patients with COPD, using medical records and artificial intelligence

Acronym
BigCOPData

Study objectives
Chronic obstructive pulmonary disease (COPD) is the third leading cause of death in the World since 2003. Many people suffer from this disease or its complications for many years and die prematurely. In the European Union, the total direct costs of respiratory diseases are estimated to be around 6% of the total healthcare budget, with COPD accounting for 56% (38.6 billion Euros) of the costs of respiratory diseases.
In the natural history of COPD, many patients may experience acute exacerbations (AECOPD) that are described as episodes of sustained worsening of the respiratory symptoms that result in additional therapy. These episodes of exacerbation that often require been seen in the emergency department and/or a hospital admission are associated with significant morbidity and mortality; they are responsible for a significant portion of the economic burden of the disease too. The pharmacological approach used in the management of AECOPD (inhaled bronchodilators, corticosteroids, and antibiotics), has the objective to minimize the negative impact of the current exacerbation and to prevent subsequent events.
Despite the collaborative effort between the European Respiratory Society, the American Thoracic Society, and others to provide clinical recommendations for the prevention of AECOPD, there is still a considerable number of patients that are prone to suffer from recurrent exacerbations and to experience a more severe impairment in health status.
Based on all the above, we aim to identify the factors potentially associated with hospital admission in patients with AECOPD in English, French, German, and Spanish, speaking countries, and to develop a predictive model that predicts the risk of hospitalization in this group of patients, by using artificial intelligence. In this study we propose to take advantage of SAVANA, a new clinical platform, created in the context of the era of electronic medical records (EMRs), to analyse the information included in the electronic medical files (i.e., big data). This clinical platform is a powerful free-text analysis engine, capable of meaningfully interpreting the contents of the EMRs, regardless of the management system in which they operate. In this context, this machine learning analytical method can be used to build a flexible, customized and automated predictive model using the information available in EMRs.

Primary objective:
To identify factors associated with hospital admission in a population of patients hospitalized for an exacerbation of COPD, and to develop a predictive hospital admission model, using EMRs and artificial intelligence

Secondary objectives:
1. To describe the clinical characteristics of COPD patients that require hospital admission
2. To identify the comorbidities associated with hospitalized COPD patients, presented per sex

(cardiovascular disease, anxiety, depression, gastroesophageal reflux, etc)
3. To identify and characterise the hospitalizations associated with increased eosinophil blood counts
4. To explore the relationship between hospitalization and inflammatory parameters such as white cell counts, neutrophil count, C-reactive protein (CRP), etc
5. To identify the clinical phenotype of patients with COPD that exacerbate and require hospital admissions
6. To explore the relationship between low adherence to treatment recommendations and hospital admission
7. To determine whether there is a relationship between hospitalization and a change of treatment in the previous 6 weeks
8. To assess stratification risk of patients, using a baseline variable (GesEPOC, the Dyspnoea, Eosinopenia, Consolidation, Acidemia and Atrial Fibrillation [DECAF] Score, or another multicomponent index)
9. To explore whether there are biologic biomarkers (different to eosinophil count) that might predict hospitalization and/or rehospitalizations due to COPD exacerbations

Ethics approval required
Old ethics approval format

Ethics approval(s)
Approved 11/04/2019, Drug Research Ethics Committee of the Princess University Hospital (CEIm La Princesa University Hospital, 62, Diego de León Street, 28006. Madrid, Spain; Tel: +34 (0) 91 520 24 76; Email: ceim.hlpr@salud.madrid.org), CEIm Act 07/19

Study design
Data-driven observational retrospective and non-interventional study using secondary data captured in EMRs

Primary study design
Observational

Secondary study design
Retrospective study

Study setting(s)
Hospital

Study type(s)
Prevention

Participant information sheet
Not applicable

Health condition(s) or problem(s) studied
Chronic obstructive pulmonary disease

Interventions
The study is retrospective, non-interventional. It’s expected to collect data from the last 5 years. The study population comprises patients who were admitted in their respective medical centres involved in the study.

The methodology data analysis is as follows:
Frequency tables will be performed for categorical variables, whereas continuous variables will be described by means of summary tables that may include the mean, standard deviation, median and range of each variable. The number of non-evaluable outcomes and of missing data will also be provided and will not be counted in the percentages. Transformations will be considered where appropriate. Unless otherwise specified, all statistical inference will be performed at the 5% significance level using 2-sided tests or 2 sided CIs.

Missing data mechanisms will be evaluated to determine appropriate methods for handling missing data when necessary (e.g. multiple imputation). A comprehensive description of the imputation procedure to ensure the transparency and reproducibility of the analysis will be provided.

This is a descriptive and hypothesis-generating study, not a confirmatory one. Therefore, other statistical models can be applied if necessary. A sensitivity analysis will be performed to deal with outliers, should it be necessary.

The last phase of the study will build a predictive model to identify those factors associated with hospital admission in a population of patients hospitalized for an exacerbation of COPD. In order to do this, the study will rely on big-data techniques that will combine advanced statistics and machine learning tools in the deep-learning spectrum. The performance of these models will be assessed in terms of precision, recall and F-score, as well as the Area Under Curve (AUC) in some cases.

Intervention Type
Other

Primary outcome measure
Given that this is a Big Data-based study, the potential number of variables that may be included is only limited to the information contained in the EMRs. All mentioned variables will be included if they are found correctly in the text. It is therefore understood that it is impossible to guarantee that all the desired variables will be included in the final study. On the other hand, this technology enables to create new variables, which can neither be described in advance.

The following variables will be extracted to meet the objectives of the study:
1. Age
2. Sex
3. Smoking status: current smoker, ex-smoker
3.1. Use of E-cigarettes, iQOS
3.2. Pack-years index
4. History of alcohol and/or drug abuse
5. Exacerbation history: number of exacerbations in the previous 12 months
6. Previous hospital admissions
7. Symptoms on admission: dyspnoea, cough, sputum, chest tightness, or wheezing
8. Clinical phenotypes
8.1. Chronic bronchitis
8.2. Emphysema
8.3. Bronchiectasis
8.4. Asthma-COPD overlap (ACO)
8.5.Frequent exacerbator
9. Pre-existing asthma

10. GOLD stage
11. Airflow obstruction
11.1. FVC
11.2. FEV1
11.3. FEV1/FVC ratio
12. mMRC dyspnea grade, if available
13. COPD Assessment Test
14. Influenza vaccination in the previous year
15. Previous pneumococcal vaccination
16. Previous microbiological isolation in sputum
17. Home oxygen therapy
18. Non-invasive mechanical ventilation (at home)
19. Mechanical ventilation (invasive and/or non-invasive) during hospital stay
20. Medication upon hospital admission, during hospitalization and hospital discharge
20.1. Inhaled corticosteroids (ICS) + LABA + LAMA
20.2. LABA + LAMA
20.3. LABA + ICS
20.4. LAMA + ICS
20.5. LAMA
20.6. LABA
20.7. ICS
20.8. Theophylline
20.9. Roflumilast
20.10. SABA / SAMA
20.11. Systemic corticosteroids
20.12. Mucolytics
20.13. Macrolides
21. Dose of systemic corticosteroids administered during hospital stay
22. Nebulized antibiotic therapy
23. Number of COPD exacerbations requiring hospitalization in the previous 12 months.
24. Number of COPD exacerbations requiring ER visits in the previous 12 months
26. Number of COPD exacerbations seen in Primary Care in the previous 12 months.
27. Blood test at hospitalization admission and sequentially during hospitalization:
27.1. Leucocytes
27.2. Neutrophils (absolute value and %)
27.3. Eosinophils (absolute value and %)
27.4. Basophils (absolute number and %)
27.5. Platelets
27.6. Haemoglobin
27.7. Fibrinogen
27.8. Urea
27.9. CRP
27.10. D-dimer
27.11. Pro-BNP-NT
27.12. Troponin
27.13. Alpha-1 antitrypsin
28. COPD-specific comorbidity test (COTE)
29. DECAF score
30. Associated comorbidities: hypertension, gastroesophageal reflux, diabetes mellitus, CV disease, skeletal muscle dysfunction, metabolic syndrome, osteoporosis, depression, anxiety and lung cancer, and other
31. Blood gas analysis, partial pressure of oxygen in arterial blood (PaO2) at hospital admission

and sequentially during hospitalization, partial pressure of carbon dioxide in arterial blood (PaCO2), pH, etc.
32. Length of hospital stay (days)
33. Ward location at hospital: respiratory unit, internal medicine unit, intensive care unit, etc.
34. Discharge location: home, home health care, nursing home, rehabilitation center, short-term hospital, other
35. Mortality during index admission
36. Hospital readmission within 30- and 90-days post-discharge

A complete and detailed guidance on the evaluation of the variables and outcomes are presented in the SAP.

Secondary outcome measures
There are no secondary outcome measures

Overall study start date
24/04/2019

Completion date
31/12/2020

Eligibility

Key inclusion criteria
1. Subjects aged ≥ 35 years old, smokers or former smokers of more than 10 pack-years
2. Had a diagnosis of COPD (a post-bronchodilator ratio forced expiratory volume in the first second [FEV1] / forced vital capacity [FVC] < 0.70, and the presence of respiratory symptoms such as cough, sputum, and dyspnoea)
3. Admitted for ''respiratory disease'' [respiratory infection or pleural effusion (OR) respiratory failure (OR) right/left heart failure (OR) chronic bronchitis (OR) bronchospasms (AND) [historical diagnosis of COPD (OR) a documented FEV1/FVC < 0.70 in the absence of other obstructive diseases, such as asthma or bronchiolitis]

Participant type(s)
Patient

Age group
Adult

Sex
Both

Target number of participants
2,500,000 patients approx

Key exclusion criteria
Patients with a specific diagnosis upon admission of pulmonary oedema, pneumonia, radiological infiltration, pulmonary embolism, pneumothorax, rib fractures, aspiration, or any other associated respiratory or of non-respiratory condition, such as major cardiopathy with chronic heart failure, extended neoplasia, liver or kidney failure.

Date of first enrolment
01/07/2019

Date of final enrolment
30/09/2020

Locations

Countries of recruitment
Austria

Belgium

England

France

Germany

Luxembourg

Spain

Switzerland

United Kingdom

Study participating centre
University Hospital Queen Elizabeth
Mindelsohn Way
Edgbaston
Birmingham
United Kingdom
B15 2GW

Sponsor information

Organisation
SEPAR (Spanish Society Pneumology and Thoracic Surgery)

Sponsor details
108, Provença Street, Bajos 2
Barcelona
Spain

08029
+34 (0)934878565
lcampos@separ.es

Sponsor type
Other

Website
https://www.separ.es

Funder(s)

Funder type
Government

Funder Name
Horizon 2020

Alternative Name(s)
EU Framework Programme for Research and Innovation, Horizon 2020 - Research and Innovation Framework Programme, European Union Framework Programme for Research and Innovation

Funding Body Type
Government organisation

Funding Body Subtype
National government

Location

Results and Publications

Publication and dissemination plan
Final results of the study will be disseminated in the form of a manuscript/s in the peer-reviewed literature. In addition, where relevant, data from potential interim analyses will be presented at (a) relevant congress(es).

Intention to publish date
01/01/2021

Individual participant data (IPD) sharing plan
The datasets generated and/or analysed during the current study during this study will be included in the subsequent results publication.

IPD sharing plan summary
Other